Amendments to the Claims:

This listing of claims replaces all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Previously amended) A system to provide finer grain control in optimizing multiple workloads across multiple servers, comprising:

a plurality of servers to be utilized by multiple workloads;

a plurality of virtual machines at each of the plurality of servers, wherein the plurality of virtual machines at each of the plurality of servers each serve a different one of the multiple workloads; and

resource management logic to distribute server resources to each of the plurality of virtual machines according to current and predicted resource needs of each of the multiple workloads utilizing the server resources,

whereby, each of the multiple workloads are distributed across the plurality of servers, wherein fractions of each of the multiple workloads are handled by the plurality of virtual machines,

whereby, the fractions of each of the multiple workloads handled by each of the virtual machines can be dynamically adjusted to provide for optimization of the server resources utilized by the multiple workloads across the multiple servers.

2. Canceled.

3. (Previously amended) The system of claim 1 wherein the server resources comprise percentage of CPU, percentage of network bandwidth, disk resources and memory resources.

4. (Original) The system of claim 1 wherein the finer grain control is achieved through recognizing when one of the plurality of servers is overloaded and shifting work to another of the plurality of servers which is not overloaded.

5. (Previously amended) The system of claim 1 wherein the fractions of the multiple workloads being handled by the plurality of virtual machines can be dynamically adjusted in response to workload changes at the plurality of servers, wherein the dynamic adjustment provides for maintaining an optimum utilization level of the server resources utilized by the multiple workloads distributed across the plurality of servers.

6. (Original) The system of claim 5 wherein the optimum utilization level can be configured automatically via server management software or manually by a user with administrative privileges.

7. (Previously amended) The system of claim 1 wherein the workloads are each distributed over a subset of the plurality of virtual machines.

8. (Previously amended) The system of claim 7 wherein each VM in the subset of the plurality of virtual machines exists at a separate one of the plurality of servers.

9. (Previously amended) The system of claim 8 wherein the workload distribution comprises distributing the work according to resources available to each of the virtual machines within the subset.

10. (Previously amended) The system of claim 1 further comprises at least one global resource allocator to monitor resource distribution between the plurality of virtual machines.

11. (Original) The system of claim 10 further comprises at least one load balancer to measure the current offered load.

12. (Previously amended) The system of claim 11 wherein the global resource allocator determines how to distribute the resources between the plurality of virtual machines, according to the measurements received from the at least one load balancer.

13. (Previously amended) The system of claim 12 wherein each of the plurality of servers includes a local resource control agent to receive and implement instructions from the global resource allocator describing how the resources are to be distributed between the virtual machines located at each of the plurality of servers.

14. (Currently amended) A server optimization device ~~, for providing finer grain control in a virtual machine based hosting architecture~~, comprising:

~~at least one load balancer component to identify resource requirements for multiple different workloads in the VM based hosting architecture;~~

a load balancer associated with a respective customer workload needing at least one workload server and for providing offered workload messages to a provider of workload servers;

a global resource allocator (GRA) for inclusion in said provider of workload servers, and for receiving said offered workload messages and assigning an optimum matching of combinations of whole integer numbers of workload servers and fractional virtual workload servers that the GRA controls to each of the respective customer workloads ~~partitioning component to assign virtual machines from multiple server machines to a workload~~ according to ~~the~~ identified resource requirements~~; and~~

~~the global resource allocator partitioning component to assign resources at each of the multiple server machines to the assigned virtual machines according to the identified resource requirements~~.

15. (Currently amended) The server optimization device of claim 14 ~~further comprises the global resource allocator partitioning component reassigning~~ wherein the fractional virtual ~~machines~~ workload servers are reassigned according to changes in the identified resource requirements of the workload assigned ~~to each virtual machine~~.

16. (Currently amended) The server optimization device of claim 14 further comprises a ~~plurality of~~ resource ~~allocator components at~~ control agent associated with each of the ~~multiple~~

~~server machines~~ whole integer number of workload servers, ~~wherein the plurality of resource allocators are responsible~~ for creating and assigning ~~virtual machines~~ fractional virtual workload servers to workloads in response to instructions received from the global resource allocator ~~partitioning component~~.

17-19. Canceled.

20. (Currently amended) The server optimization device of claim 14 wherein the identified resource requirements ~~resources~~ comprise the percentage of CPU, percentage of network bandwidth, disk resources and memory resources that are needed by a workload.

21-23. Canceled.

24. (Previously amended) A method for improving server utilization levels, comprising:

identifying a current offered load of each of a plurality of customer applications; analyzing the identified current offered load and generating a prediction as to what resources will be needed by each of the plurality of customer applications;

identifying all virtual machines associated with each of the plurality of customer applications; and

allocating, according to the generated prediction, the resources needed by each of the plurality of customer applications to the virtual machines associated with each of the plurality of customer applications,

whereby, the offered load associated each of the customer applications is distributed over a plurality of the identified virtual machines, and the virtual machines over which each of the customer applications offered load is distributed, reside on separate machines.

25. (Currently amended) A server optimization ~~means~~ device ~~, for providing finer grain control in a virtual machine (VM) based hosting architecture~~, comprising:

a ~~means~~ machine for identifying resource requirements for multiple different workloads in the VM virtual machine based hosting architecture;

a ~~means~~ machine for assigning virtual machines from multiple server machines to a workload according to the identified resource requirements; and

a ~~means~~ machine for assigning resources at each of the multiple server machines to the assigned virtual machines according to the identified resource requirements.

26. (Currently amended) The server optimization ~~means~~ device of claim 25 further comprises a ~~means~~ machine for reassigning the virtual machines according changes in the identified resource requirements.

27. (Currently amended) The server optimization ~~means~~ device of claim 25 further comprises a ~~means~~ mechanism for creating virtual machines and assigning virtual machines to workloads in response to instructions received from the global resource allocator partitioning component.

28. (Previously amended) A computer program product for use with a computer hosting architecture, for providing finer grain control in a virtual machine based hosting architecture, comprising:

a computer-readable

medium means, provided on the computer-readable medium, for identifying resource requirements for multiple different workloads in the virtual machine based hosting architecture;

means, provided on the computer-readable medium, for assigning virtual machines from multiple server machines to a workload according to the identified resource requirements; and

means, provided on the computer-readable medium, for assigning resources at each of the multiple server machines to the assigned virtual machines according to the identified resource requirements.